A Method for Convex Black-Box Integer Global Optimization

Jeffrey Larson, Sven Leyffer, Prashant Palkar, Stefan M. Wild

Argonne National Laboratory

July 15, 2019

Problem formulation

Derivative-Free Optimization with Unrelaxable Integers

$$\underset{x}{\text{minimize}} \;\; f(S(x)) \quad \text{subject to } x \in X \subset \mathbb{Z}^p \times \mathbb{R}^q$$

1. Evaluation involves $S(x)$ numerical simulation (computationally expensive)
   - derivatives $\nabla_x S$ unavailable or expensive to compute
   - single evaluation of $S(x)$ can take minutes/hours/days
2. Unrelaxable integers, e.g. # receiver panels
   - Unrelaxable: simulation cannot run at fractional values!

Background

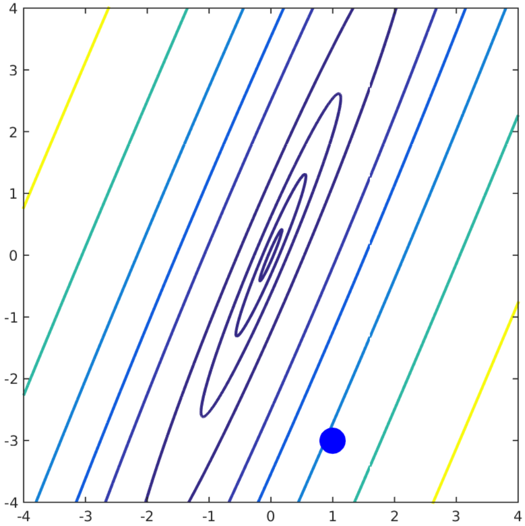
Some applications

- Design of concentrating solar power plants (*Pascal et al, 2011*)
- Performance tuning codes on high-performance computers (*Balaprakash et al, 2014*) etc.

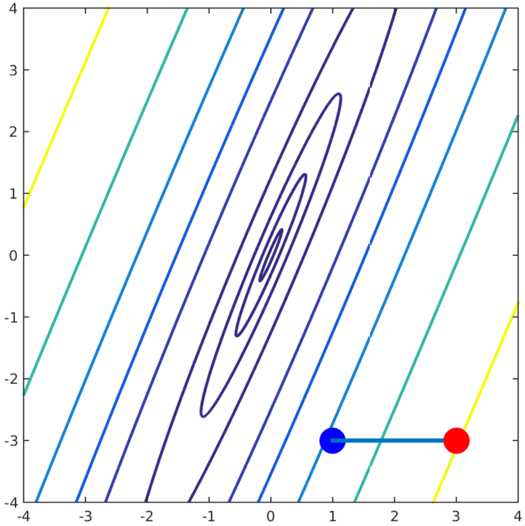
Addressing integer variables

- heuristic approaches: rounding integer variables (*Mueller et al, 2013*)
- patten-search methods (*Abramson et al, 2008; Audet et al, 2001*)
- definitions of minimizers and deficiencies (*Newby and Ali, 2015*)
- primitive directions and nonmonotone line searches with integer variables (*Liuzzi et al, 2018*)
- no outer approximation ($\nabla_x S$ is unavailable)
- no branch-and-bound (unrelaxable integer constraints)

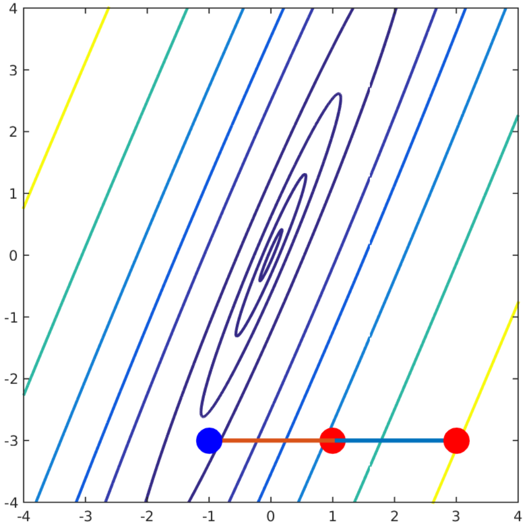
Pattern search



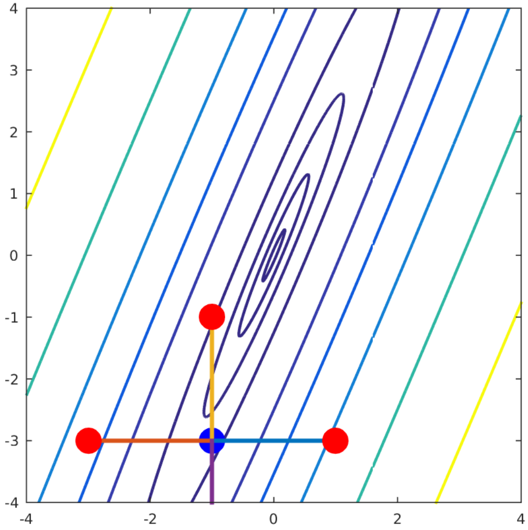
Pattern search



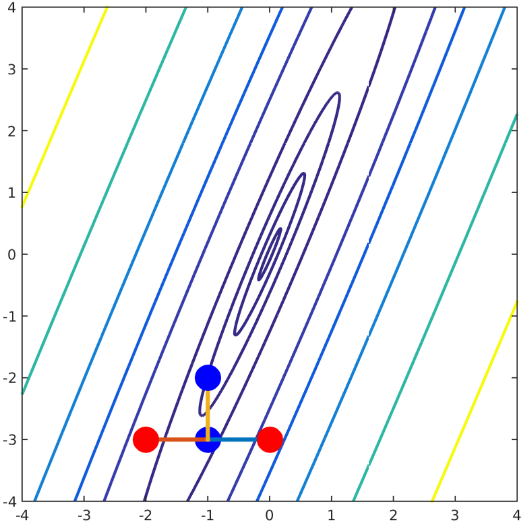
Pattern search



Pattern search



Pattern search



Pattern search

Proposed by Audet & Dennis (2001):

- User-defined discrete neighborhood

Pattern search

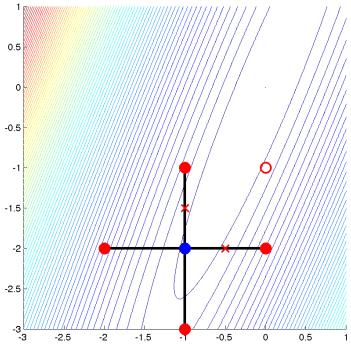
Proposed by Audet & Dennis (2001):

- User-defined discrete neighborhood
- Declare "mesh-isolated minimizer" if no local improvement

Pattern search

Proposed by Audet & Dennis (2001):

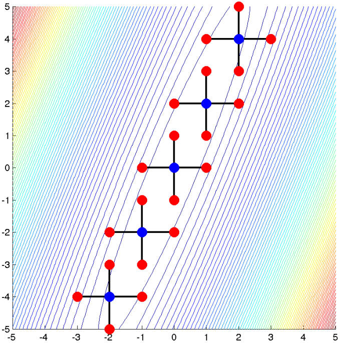
- User-defined discrete neighborhood
- Declare "mesh-isolated minimizer" if no local improvement



Pattern search

Proposed by Audet & Dennis (2001):

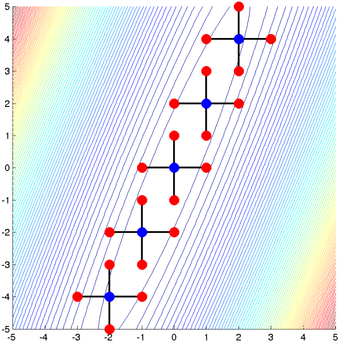
- User-defined discrete neighborhood
- Declare "mesh-isolated minimizer" if no local improvement



Pattern search

Proposed by Audet & Dennis (2001):

- User-defined discrete neighborhood
- Declare "mesh-isolated minimizer" if no local improvement



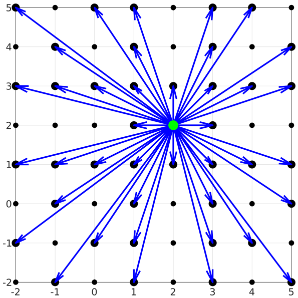
- Any $(y_1, y_2) \in \mathbb{Z}^2$ with $2y_1 = y_2$ is optimal

Discussion

Question

Can we guarantee a global minimizer of a convex $f(x)$ when x is integer?

Primitive directions



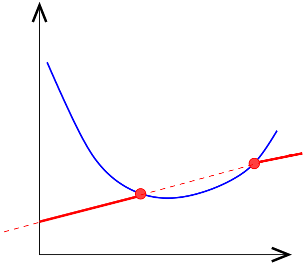
	$n=2$		$n=3$		$n=4$		$n=5$	
k	$\|\Omega\|$	#	$\|\Omega\|$	#	$\|\Omega\|$	#	$\|\Omega\|$	#
1	9	8	27	26	81	80	243	242
2	25	16	125	98	625	544	3,125	2,882
3	49	32	343	290	2,403	2,240	16,807	16,322
4	81	48	729	578	6,561	5,856	59,049	55,682

Table: Number of primitive directions, $\# = |\mathcal{N}(x_c, 1)|$, that emanate from the origin x_c of the domain $\Omega = [-k, k]^n \cap \mathbb{Z}^n$ and that correspond to points in Ω.

1D

$$\underset{x}{\text{minimize}}\ f(x), \quad \text{subject to } x \in \mathbb{Z}^n$$

and assume $f(x)$ convex

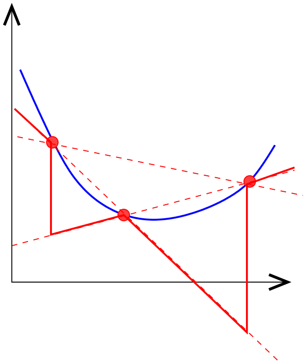


. . . underestimator for convex, integer DFO!

1D

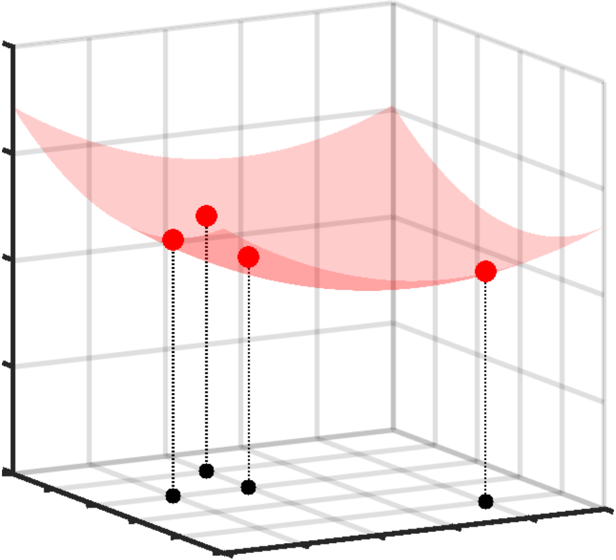
$$\underset{x}{\text{minimize}} \;\; f(x), \quad \text{subject to } x \in \mathbb{Z}^n$$

and assume $f(x)$ convex

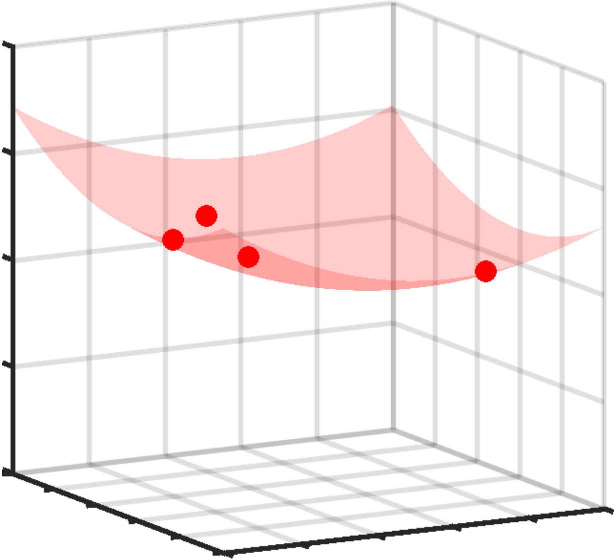


. . . underestimator for convex, integer DFO!

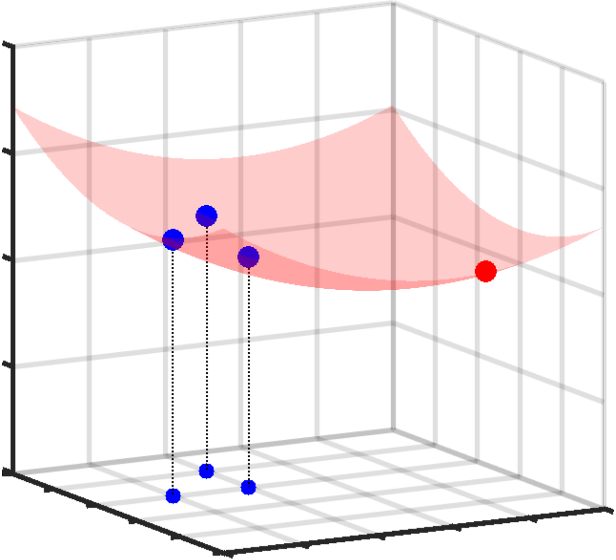
2D



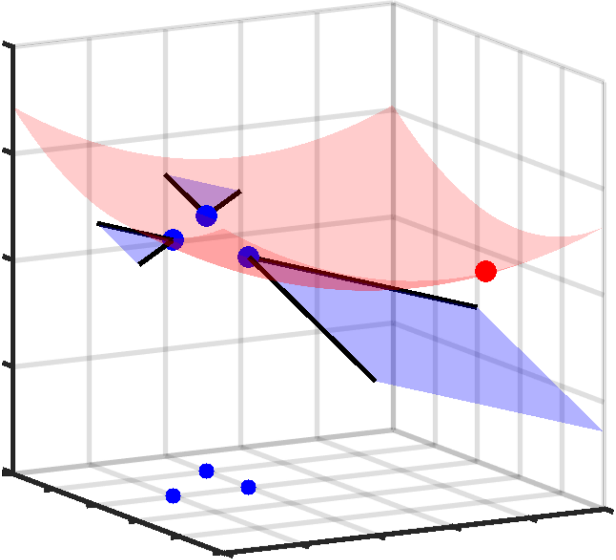
2D



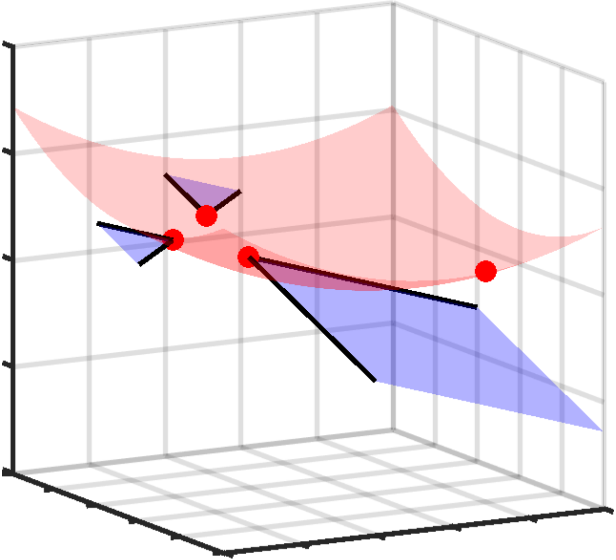
2D



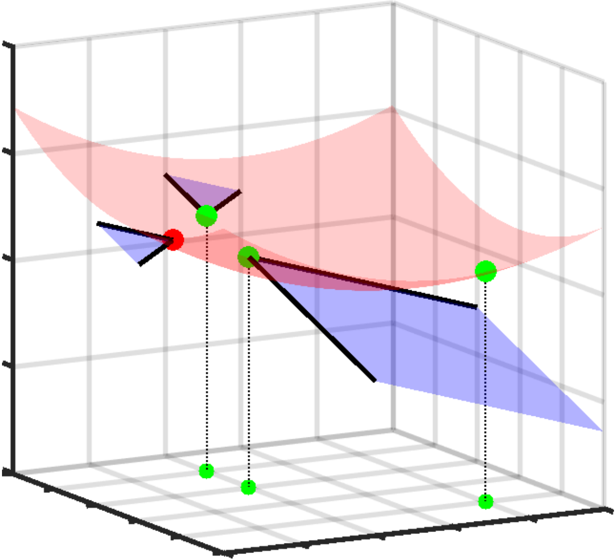
2D



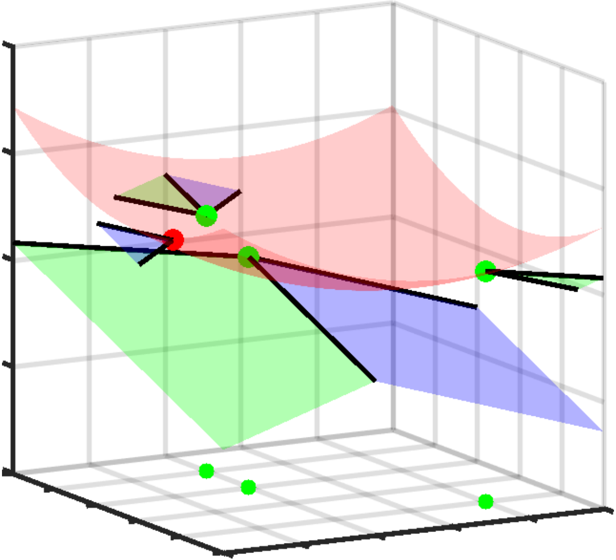
2D



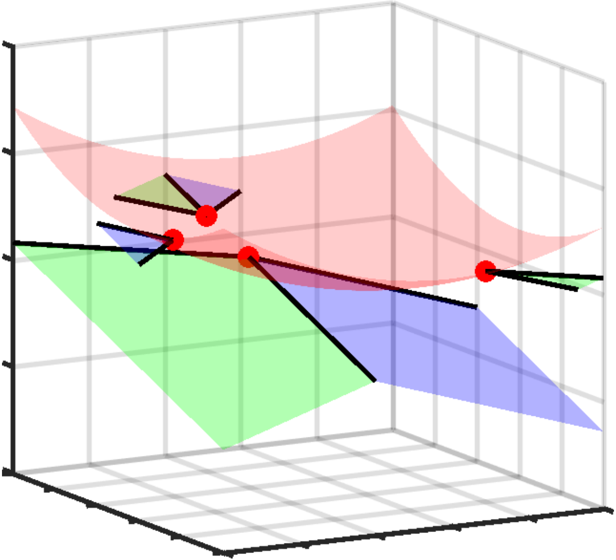
2D



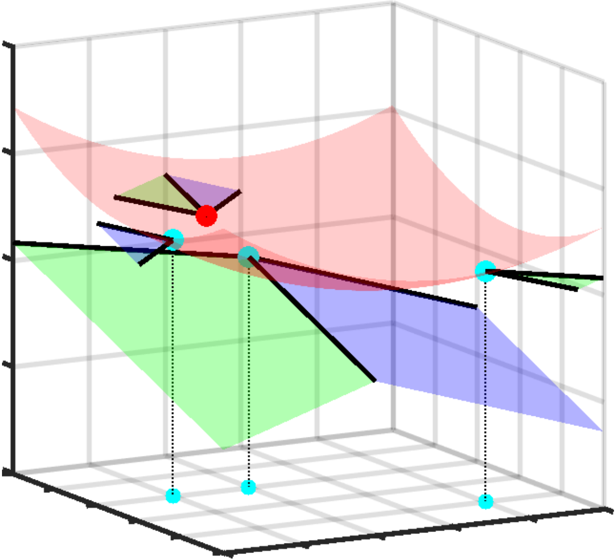
2D



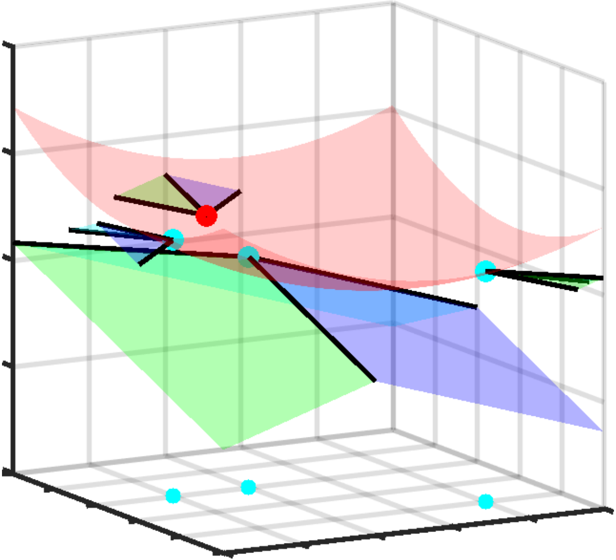
2D



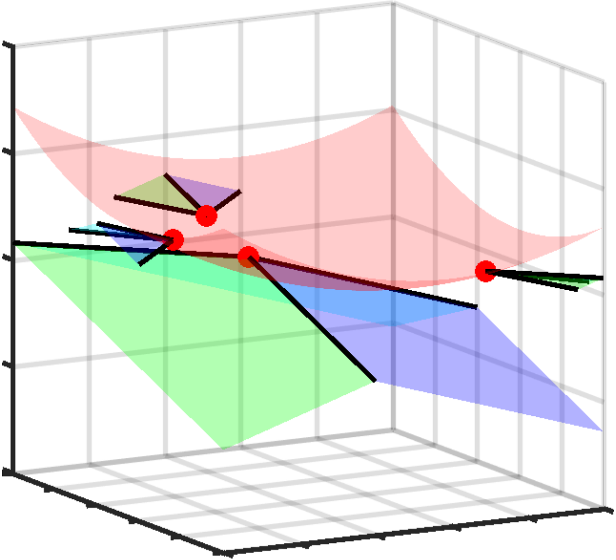
2D



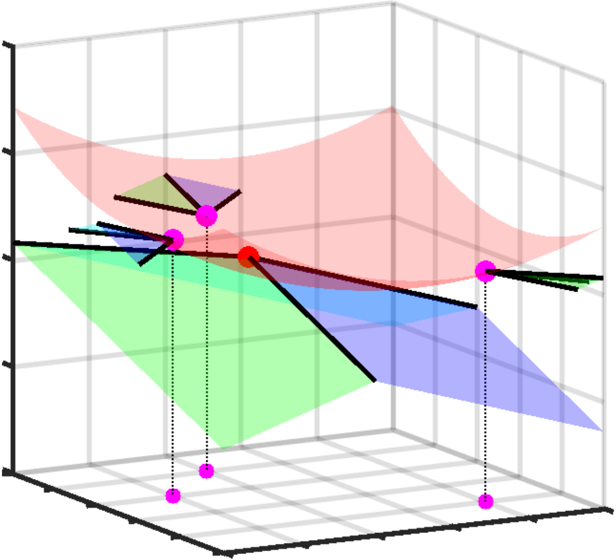
2D



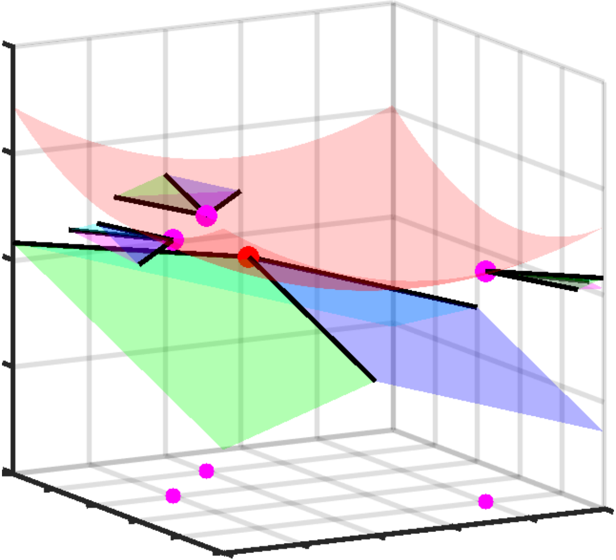
2D



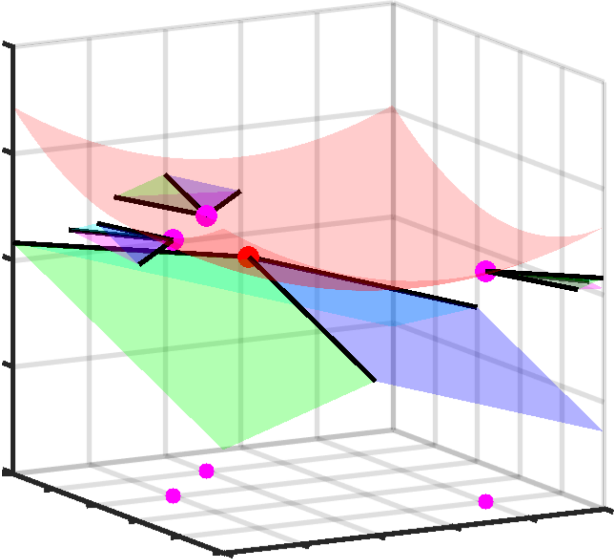
2D



2D



2D



Underestimating f

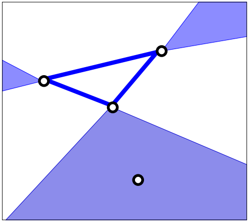
Formulate piecewise underestimator as MILP

- Interpolation points: $X := \{x^i \in \mathbb{Z}^n\}$, $|X| \geq n+1$
- Function values: $f^i := f(x^i)$ for $x^i \in X$
- $\mathbf{i} := (i_1, \ldots, i_{n+1})$ multi-index for $n+1$ distinct $i_j \in \mathbf{i}$ with $1 \leq i_1 < \ldots, i_{n+1} \leq |X|$

Interpolation Cuts

For $X^{\mathbf{i}} := \{x^{i_j} : i_j \in \mathbf{i}\}$ obtain cut $(c^{\mathbf{i}})^T x + b^{\mathbf{i}}$
...only valid in cones ...by solving linear system:

$$\begin{bmatrix} X^{\mathbf{i}} & e \end{bmatrix} \begin{bmatrix} c^{\mathbf{i}} \\ b^{\mathbf{i}} \end{bmatrix} = f^{\mathbf{i}},$$



Underestimating f

Lemma: Underestimating Property

$f(x)$ convex and $X^{\mathbf{i}} = \{x^{i_1}, \ldots, x^{i_{n+1}}\}$ poised, then if follows that

$$f(x) \geq (c^{\mathbf{i}})^T x + b^{\mathbf{i}}, \qquad \forall\, x \in U^{\mathbf{i}} := \bigcup_{j=1}^{n+1} \mathrm{cone}\left(x^{i_j} - X^{\mathbf{i}}\right),$$

where $\mathrm{cone}\left(x^{i_j} - X^{\mathbf{i}}\right)$ is the cone with vertex $x^{i_j} \in X^{\mathbf{i}}$ & rays $x^{i_j} - x^{i_l}$:

Solving the subproblem - MILP formulation

Modeling membership in cones

- Binary $z^{i_j} = 1$ if and only if $x \in \mathrm{cone}\left(x^{i_j} - X^{\mathbf{i}}\right)$, for $i_j \in \mathbf{i}$

Solving the subproblem - MILP formulation

Modeling membership in cones

- Binary $z^{i_j} = 1$ if and only if $x \in \operatorname{cone}\left(x^{i_j} - X^{\mathbf{i}}\right)$, for $i_j \in \mathbf{i}$
- Cut $\eta \geq (c^{\mathbf{i}})^T x + b^{\mathbf{i}} - M_{\mathbf{i}}(1 - \sum_{j=1}^{n+1} z^{i_j})$ for big-$M_{\mathbf{i}} > 0$

Solving the subproblem - MILP formulation

Modeling membership in cones

- Binary $z^{i_j} = 1$ if and only if $x \in \mathrm{cone}\left(x^{i_j} - X^{\mathbf{i}}\right)$, for $i_j \in \mathbf{i}$
- Cut $\eta \geq (c^{\mathbf{i}})^T x + b^{\mathbf{i}} - M_{\mathbf{i}}(1 - \sum_{j=1}^{n+1} z^{i_j})$ for big-$M_{\mathbf{i}} > 0$
- SOS-1 constraint: at most one cone, $z^{i_j} \leq 1$, active

Solving the subproblem - MILP formulation

Modeling membership in cones

- Binary $z^{i_j} = 1$ if and only if $x \in \mathrm{cone}\left(x^{i_j} - X^{\mathbf{i}}\right)$, for $i_j \in \mathbf{i}$
- Cut $\eta \geq (c^{\mathbf{i}})^T x + b^{\mathbf{i}} - M_{\mathbf{i}}(1 - \sum_{j=1}^{n+1} z^{i_j})$ for big-$M_{\mathbf{i}} > 0$
- SOS-1 constraint: at most one cone, $z^{i_j} \leq 1$, active
- Any point x is linear combination of extreme rays ($W(X)$ set of all poised subsets)

$$x = x^{i_j} + \sum_{\substack{l=1,\\ l \neq j}}^{n+1} \lambda_l^{i_j} (x^{i_j} - x^{i_l}), \quad \forall i_j \in \mathbf{i},\ \forall \mathbf{i} \in W(X)$$

Solving the subproblem - MILP formulation

Modeling membership in cones

- Binary $z^{i_j} = 1$ if and only if $x \in \operatorname{cone}\left(x^{i_j} - X^{\mathbf{i}}\right)$, for $i_j \in \mathbf{i}$
- Cut $\eta \geq (c^{\mathbf{i}})^T x + b^{\mathbf{i}} - M_{\mathbf{i}}(1 - \sum_{j=1}^{n+1} z^{i_j})$ for big-$M_{\mathbf{i}} > 0$
- SOS-1 constraint: at most one cone, $z^{i_j} \leq 1$, active
- Any point x is linear combination of extreme rays ($W(X)$ set of all poised subsets)

$$x = x^{i_j} + \sum_{\substack{l=1,\\ l\neq j}}^{n+1} \lambda_l^{i_j} (x^{i_j} - x^{i_l}), \quad \forall i_j \in \mathbf{i},\ \forall \mathbf{i} \in W(X)$$

- Indicate $x \in \operatorname{cone}\left(x^{i_j} - X^{\mathbf{i}}\right)$ by making $\lambda_l^{i_j} \geq -M_\lambda(1 - z^{i_j})$

Solving the subproblem - MILP formulation

Modeling membership in cones

- Binary $z^{i_j} = 1$ if and only if $x \in \operatorname{cone}\left(x^{i_j} - X^{\mathbf{i}}\right)$, for $i_j \in \mathbf{i}$
- Cut $\eta \geq (c^{\mathbf{i}})^T x + b^{\mathbf{i}} - M_{\mathbf{i}}(1 - \sum_{j=1}^{n+1} z^{i_j})$ for big-$M_{\mathbf{i}} > 0$
- SOS-1 constraint: at most one cone, $z^{i_j} \leq 1$, active
- Any point x is linear combination of extreme rays ($W(X)$ set of all poised subsets)

$$x = x^{i_j} + \sum_{\substack{l=1,\\ l \neq j}}^{n+1} \lambda_l^{i_j} (x^{i_j} - x^{i_l}), \quad \forall i_j \in \mathbf{i},\ \forall \mathbf{i} \in W(X)$$

- Indicate $x \in \operatorname{cone}\left(x^{i_j} - X^{\mathbf{i}}\right)$ by making $\lambda_l^{i_j} \geq -M_\lambda(1 - z^{i_j})$

... models $z^{i_j} = 1 \Rightarrow x \in \operatorname{cone}\left(x^{i_j} - X^{\mathbf{i}}\right)$ for $i_j \in \mathbf{i}$ ... reverse harder

An alternative master problem

Challenges of MILP Master model

- MILP model exponential in number of interpolation points
- MILP representation is very weak: uses big-M and tiny-ϵ

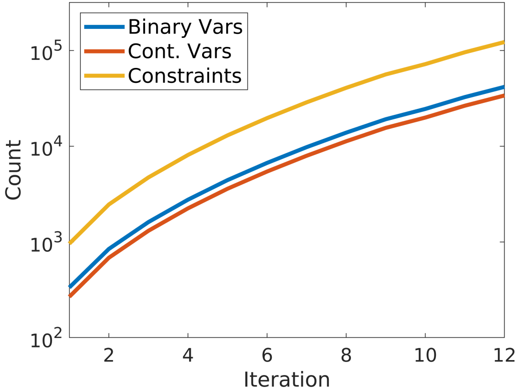
$\Rightarrow$ Commercial solvers cannot solve large instances

An alternative master problem

Challenges of MILP Master model

- MILP model exponential in number of interpolation points
- MILP representation is very weak: uses big-M and tiny-ϵ

$\Rightarrow$ Commercial solvers cannot solve large instances



First 12 instances of MIP model while minimizing the convex quadratic `abhi` on $\Omega = [-2,2]^3 \cap \mathbb{Z}^3$

An alternative master problem

Replacing CPLEX Solve by Look-Up Table

- Key idea: work in space of original integers, $x \in \mathbb{Z}^n$ (no additional variables or constraints)
- Replace MILP by look-up-table of underestimator
- Update look-up-table when new points (and therefore new cuts) are available

An alternative master problem

Replacing CPLEX Solve by Look-Up Table

- Key idea: work in space of original integers, $x \in \mathbb{Z}^n$ (no additional variables or constraints)
- Replace MILP by look-up-table of underestimator
- Update look-up-table when new points (and therefore new cuts) are available

Dense/small linear algebra solves $\Rightarrow$ Fast

An alternative master problem

Replacing CPLEX Solve by Look-Up Table

- Key idea: work in space of original integers, $x \in \mathbb{Z}^n$ (no additional variables or constraints)
- Replace MILP by look-up-table of underestimator
- Update look-up-table when new points (and therefore new cuts) are available

Dense/small linear algebra solves $\Rightarrow$ Fast . . . but not fast enough

Algorithm 1: Look up algorithm

input: Lower bound η for each point in Ω; Points X with $|X| \geq n+1$; $\bar{f} = \min_{x_i \in X} f(x_i)$

Algorithm 1: Look up algorithm

input: Lower bound η for each point in Ω; Points X with $|X| \geq n+1$; $\bar{f} = \min_{x_i \in X} f(x_i)$

while $\bar{f} > \min \eta$ **do**

Algorithm 1: Look up algorithm

input: Lower bound η for each point in Ω; Points X with $|X| \geq n+1$; $\bar{f} = \min_{x_i \in X} f(x_i)$

while $\bar{f} > \min \eta$ **do**

  $C = \{$subsets of $n+1$ points in $X\}$

Algorithm 1: Look up algorithm

input: Lower bound η for each point in Ω; Points X with $|X| \geq n+1$; $\bar{f} = \min_{x_i \in X} f(x_i)$

while $\bar{f} > \min \eta$ **do**

  $C = \{\text{subsets of } n+1 \text{ points in } X\}$

  for $\mathbf{i} \in C$ **do**

    if $X^{\mathbf{i}}$ *is poised* **then**

Algorithm 1: Look up algorithm

input: Lower bound η for each point in Ω; Points X with $|X| \geq n+1$; $\bar{f} = \min_{x_i \in X} f(x_i)$

while $\bar{f} > \min \eta$ **do**

  $C = \{$subsets of $n+1$ points in $X\}$

  for $\mathbf{i} \in C$ **do**

    if $X^{\mathbf{i}}$ *is poised* **then**

      Find points in Ω in one of the cones of $X^{\mathbf{i}}$

Algorithm 1: Look up algorithm

input: Lower bound η for each point in Ω; Points X with $|X| \geq n+1$; $\bar{f} = \min\limits_{x_i \in X} f(x_i)$

while $\bar{f} > \min \eta$ **do**

  $C = \{$subsets of $n+1$ points in $X\}$

  for $\mathbf{i} \in C$ **do**

    if $X^{\mathbf{i}}$ *is poised* **then**

      Find points in Ω in one of the cones of $X^{\mathbf{i}}$

      Update η

Algorithm 1: Look up algorithm

input: Lower bound η for each point in Ω; Points X with $|X| \geq n+1$; $\bar{f} = \min_{x_i \in X} f(x_i)$

while $\bar{f} > \min \eta$ **do**

  $C = \{\text{subsets of } n+1 \text{ points in } X\}$

  for $\mathbf{i} \in C$ **do**

    if $X^{\mathbf{i}}$ *is poised* **then**

      Find points in Ω in one of the cones of $X^{\mathbf{i}}$

      Update η

  Add $x^k = \arg\min_{x \in \Omega} \eta$ and update $\bar{f}$

Algorithm 1: Look up algorithm

input: Lower bound η for each point in Ω; Points X with $|X| \geq n+1$; $\bar{f} = \min_{x_i \in X} f(x_i)$

while $\bar{f} > \min \eta$ **do**

  $C = \{$subsets of $n+1$ points in $X\}$

  for $\mathbf{i} \in C$ **do**

    $[Q, R] = \mathrm{qr}([e\ X^{\mathbf{i}}])$

    if $X^{\mathbf{i}}$ *is poised* **then**

      Find points in Ω in one of the cones of $X^{\mathbf{i}}$

      Update η

  Add $x^k = \arg\min_{x \in \Omega} \eta$ and update $\bar{f}$

Algorithm 1: Look up algorithm

input: Lower bound η for each point in Ω; Points X with $|X| \geq n+1$; $\bar{f} = \min_{x_i \in X} f(x_i)$

while $\bar{f} > \min \eta$ **do**

  $C = \{\text{subsets of } n+1 \text{ points in } X\}$

  for $\mathbf{i} \in C$ **do**

    $[Q, R] = \mathrm{qr}([e\ X^{\mathbf{i}}])$

    if $X^{\mathbf{i}}$ *is poised using* R **then**

      Find points in Ω in one of the cones of $X^{\mathbf{i}}$

      Update η

  Add $x^k = \arg\min_{x \in \Omega} \eta$ and update $\bar{f}$

Algorithm 1: Look up algorithm

input: Lower bound η for each point in Ω; Points X with $|X| \geq n+1$; $\bar{f} = \min_{x_i \in X} f(x_i)$

while $\bar{f} > \min \eta$ **do**

  $C = \{\text{subsets of } n+1 \text{ points in } X\}$

  for $\mathbf{i} \in C$ **do**

    $[Q, R] = \mathrm{qr}([e \; X^{\mathbf{i}}])$

    if $X^{\mathbf{i}}$ *is poised using* R **then**

      Find points in Ω in one of the cones of $X^{\mathbf{i}}$ using Q

      Update η

  Add $x^k = \arg\min_{x \in \Omega} \eta$ and update $\bar{f}$

Algorithm 1: Look up algorithm

input: Lower bound η for each point in Ω; Points X with $|X| \geq n+1$; $\bar{f} = \min_{x_i \in X} f(x_i)$

while $\bar{f} > \min \eta$ **do**

  $C = \{$subsets of $n+1$ points in $X\}$

  for $\mathbf{i} \in C$ **do**

    $[Q, R] = \text{qr}([e\ X^{\mathbf{i}}])$

    if $X^{\mathbf{i}}$ *is poised using* R **then**

      Find points in Ω in one of the cones of $X^{\mathbf{i}}$ using Q

      Update η using Q

  Add $x^k = \arg\min_{x \in \Omega} \eta$ and update $\bar{f}$

Algorithm 1: Look up algorithm

input: Lower bound η for each point in Ω; Points X with $|X| \geq n+1$; $\bar{f} = \min_{x_i \in X} f(x_i)$

while $\bar{f} > \min \eta$ **do**

  $C = \{$subsets of $n+1$ points in $X\}$

  for $\mathbf{i} \in C$ **do**

    $[Q, R] = \mathrm{qr}([e\ X^{\mathbf{i}}])$

    if $X^{\mathbf{i}}$ *is poised using* R **then**

      Find points in Ω in one of the cones of $X^{\mathbf{i}}$ using Q

      Update η using Q where $\eta < \bar{f}$

  Add $x^k = \arg\min_{x \in \Omega} \eta$ and update $\bar{f}$

Algorithm 1: Look up algorithm

input: Lower bound η for each point in Ω; Points X with $|X| \geq n+1$; $\bar{f} = \min_{x_i \in X} f(x_i)$

while $\bar{f} > \min \eta$ **do**

  $C = \{$subsets of $n+1$ points in $X\}$

  for $\mathbf{i} \in C$ **do**

    $[Q, R] = \text{qr}([e\ X^{\mathbf{i}}])$

    if $X^{\mathbf{i}}$ *is poised using* R **then**

      Find points in Ω in one of the cones of $X^{\mathbf{i}}$ using Q

      Update η using Q where $\eta < \bar{f}$

  Add $x^k = \arg\min_{x \in \Omega, \eta < \bar{f}} \eta$ and update $\bar{f}$

Algorithm 1: Look up algorithm

---

input: Lower bound η for each point in Ω; Points X with $|X| \geq n+1$; $\bar{f} = \min_{x_i \in X} f(x_i)$

while $\bar{f} > \min \eta$ **do**

- $C = \{\text{subsets of } n \text{ points in } X\} \otimes \{x^k\}$
- **for** $\mathbf{i} \in C$ **do**
  - $[Q, R] = \mathrm{qr}([e\ X^{\mathbf{i}}])$
  - **if** $X^{\mathbf{i}}$ *is poised using* R **then**
    - Find points in Ω in one of the cones of $X^{\mathbf{i}}$ using Q
    - Update η using Q where $\eta < \bar{f}$
- Add $x^k = \arg\min_{x \in \Omega, \eta < \bar{f}} \eta$ and update $\bar{f}$

Algorithm 1: Look up algorithm

---

input: Lower bound η for each point in Ω; Points X with $|X| \geq n+1$; $\bar{f} = \min_{x_i \in X} f(x_i)$

while $\bar{f} > \min \eta$ **do**

- $C = \{\text{subsets of } n+1 \text{ useful points in } X\}$
- **for** $\mathbf{i} \in C$ **do**
  - $[Q, R] = \mathrm{qr}([e\ X^{\mathbf{i}}])$
  - **if** $X^{\mathbf{i}}$ *is poised using* R **then**
    - Find points in Ω in one of the cones of $X^{\mathbf{i}}$ using Q
    - Update η using Q where $\eta < \bar{f}$
- Add $x^k = \arg\min_{x \in \Omega, \eta < \bar{f}} \eta$ and update $\bar{f}$

Algorithm 1: Look up algorithm

input: Lower bound η for each point in Ω; Points X with $|X| \geq n+1$; $\bar{f} = \min_{x_i \in X} f(x_i)$

while $\bar{f} > \min \eta$ **do**

  Generate a sufficient set C of subsets of $n+1$ points

  for $\mathbf{i} \in C$ **do**

    $[Q, R] = \mathrm{qr}([e\ X^{\mathbf{i}}])$

    if $X^{\mathbf{i}}$ *is poised using* R **then**

      Find points in Ω in one of the cones of $X^{\mathbf{i}}$ using Q

      Update η using Q where $\eta < \bar{f}$

  Add $x^k = \arg\min_{x \in \Omega, \eta < \bar{f}} \eta$ and update $\bar{f}$

Algorithm 1: Look up algorithm

input: Lower bound η for each point in Ω; Points X with $|X| \geq n+1$; $\bar{f} = \min_{x_i \in X} f(x_i)$

while $\bar{f} > \min \eta$ **do**

  Generate a sufficient set C of subsets of $n+1$ points

  for $\mathbf{i} \in C$ **do**

    $[Q, R] = \text{qr}([e\ X^{\mathbf{i}}])$

    if $X^{\mathbf{i}}$ *is poised using* R **then**

      Find points in Ω in one of the cones of $X^{\mathbf{i}}$ using Q

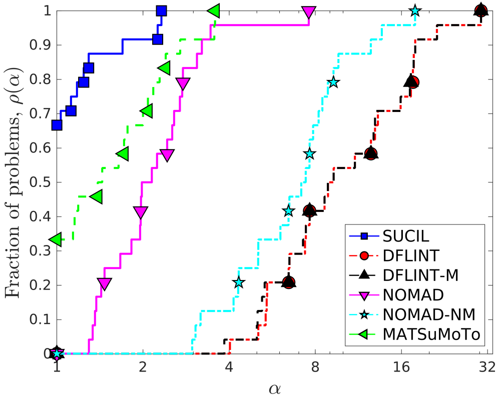
      Update η using Q where $\eta < \bar{f}$

  Add $x^k = \arg\min_{x \in \Omega, \eta < \bar{f}} \eta$ and update $\bar{f}$

Question

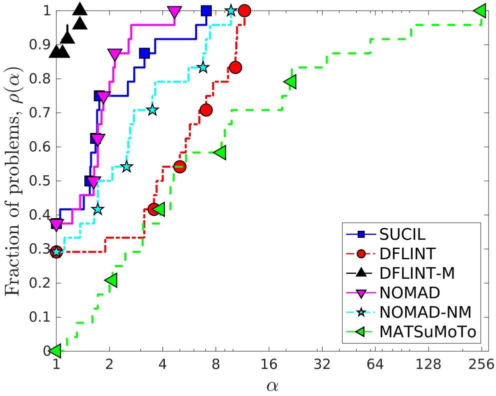
Given a set of points X on the integer lattice, is there a way to generate all subsets of size $n+1$ without another in the interior?

Benchmarking on 24 convex problems



Comparing the number of evaluations before...
each method terminates

Benchmarking on 24 convex problems



Comparing the number of evaluations before...
each method first evaluates a global minimizer

Future

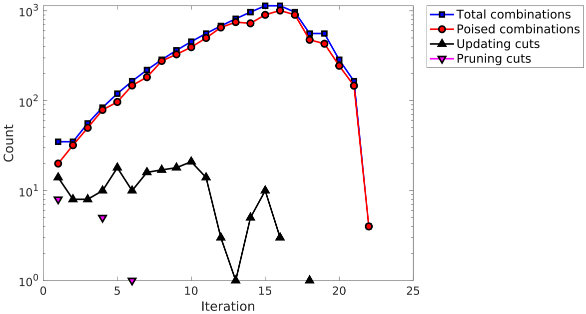
Question

Better approach for determining important cuts using information in both f- and x-space? (We aren't using convexity as much as possible.)

Future

Question

Better approach for determining important cuts using information in both f- and x-space? (We aren't using convexity as much as possible.)

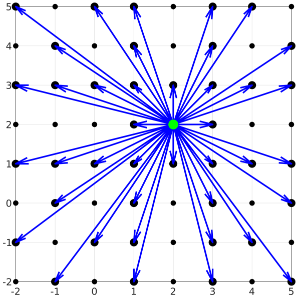


Minimizing $f(x) = \|x\|_2^2$ on $[-4, 4]^3 \bigcup \mathbb{Z}^3$.
(729 points, 578 *primitive directions* emanating from origin.)

Future

Question

Better approach for determining important cuts using information in both f- and x-space? (We aren't using convexity as much as possible.)



Future

Question

Better approach for determining important cuts using information in both f- and x-space? (We aren't using convexity as much as possible.)

Question

How to certify (local) optimality when f is nonconvex?